

High-Level Visual Object Representations Are Constrained by Position

Dwight J. Kravitz, Nikolaus Kriegeskorte and Chris I. Baker

Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA

Address correspondence to Dwight J. Kravitz. Email: kravitzd@mail.nih.gov

It is widely assumed that high-level visual object representations are position-independent (or invariant). While there is sensitivity to position in high-level object-selective cortex, position and object identity are thought to be encoded independently in the population response such that position information is available across objects and object information is available across positions. Contrary to this view, we show, with both behavior and neuroimaging, that visual object representations are position-dependent (tied to limited portions of the visual field). Behaviorally, we show that the effect of priming an object was greatly reduced with any change in position (within- or between-hemifields), indicating nonoverlapping representations of the same object across different positions. Furthermore, using neuroimaging, we show that object-selective cortex is not only highly sensitive to object position but also the ability to differentiate objects based on its response is greatly reduced across different positions, consistent with the observed behavior and the receptive field properties observed in macaque object-selective neurons. Thus, even at the population level, the object information available in response of object-selective cortex is constrained by position. We conclude that even high-level visual object representations are position-dependent.

Keywords: behavior, fMRI, object recognition, position invariance, vision

Introduction

Despite retinal position being a fundamental aspect of visual input, object recognition is widely assumed to be position-invariant or -independent (Riesenhuber and Poggio 2000; Dicarlo and Cox 2007; Hoffman and Logothetis 2009), consistent with our phenomenological experience of recognizing objects equally well across the visual field. Most accounts propose that this independence arises from visual object representations that are themselves position-independent or at least highly tolerant of position changes (Riesenhuber and Poggio 2000; Dicarlo and Cox 2007). These representations could be simple, as in a single visually responsive neuron with a large receptive fields (RFs), or complex, as in the ability to “read out” object information across positions from the population response.

At the single-unit level in macaque inferior temporal (IT) cortex, the rank order of responsiveness to complex stimuli is often maintained across the RF (DiCarlo and Maunsell 2003; Yamane et al. 2008), suggesting that position invariance may be maintained within the RF despite changes in absolute response. However, while there may be some IT neurons with RFs covering large portions of the visual field (Gross et al. 1972), RF size is heterogeneous (Op De Beeck and Vogels 2000), with some RFs covering less than 1.5 degrees (DiCarlo and Maunsell 2003). Thus, individual neurons show only limited invariance making it difficult to infer the extent to which the output of IT cortex is position-independent.

Despite the large proportion of IT neurons with small RFs, it has recently been argued that position-independence emerges at the population level (Hung et al. 2005; Schwarzlose et al. 2008). For example, in single-unit data, linear classifiers were found to be able to provide object information across changes in position as well as size (Hung et al. 2005). Similarly, human functional magnetic resonance imaging (fMRI) studies have suggested that category information can be read out across positions using the pattern of response across a region of cortex even though there is sensitivity to position (Sayres and Grill-Spector 2008; Schwarzlose et al. 2008; Williams et al. 2008; Macevoy and Epstein 2009).

While position-independent population readout is appealing, it is unclear whether a population can evidence greater position-independence than its constituent neurons (Goris and Op de Beeck 2009). In the single-unit data supporting invariant readout (Hung et al. 2005), the position shifts tested were always within the contralateral field and small enough (~4 degrees) that they would easily fit within the average RF size. Similarly, the human fMRI studies have generally only test categorization (among a small set of possible categories across within-field shifts of position [Sayres and Grill-Spector 2008; Schwarzlose et al. 2008; Williams et al. 2008]). The differences between the categories tested have also tended to be very large (e.g., faces vs. scenes [Schwarzlose et al. 2008]), all of which might have contributed to the reported position-independence (see Supplementary Item 1). Furthermore, these studies lacked supporting behavioral data, which would have more firmly established how the ability to read out relates to the behavioral output (e.g., Williams et al. 2007).

The importance of obtaining converging behavioral evidence is magnified as increasingly complex aspects of the neural response are derived via multivariate and classifier-based approaches. While there are some behavioral reports of position-independence, even across visual fields (Biederman and Cooper 1991; Fiser and Biederman 2001), the existing behavioral literature offers little clarity, with contradictory results that might reflect more the specific stimuli and tasks employed than the nature of the underlying visual representations (see Kravitz et al. 2008 for a review).

Here, we used 2 different approaches to investigate the position-dependence of visual object representations. Behaviorally, we found significant reductions in object priming with changes in position, indicating nonoverlapping visual representations. Furthermore, with fMRI, we found the ability to differentiate between objects (individuation) using the response of high-level object-selective cortex was significantly weaker across positions than within a position. Taken together, these behavioral and imaging findings show that visual object representations are position-dependent, contradicting an assumption present throughout much of the object recognition literature.

Materials and Methods

Behavioral Stimuli and Task

Participants were briefly presented (66 or 150 ms) with a “whole” or “scrambled” line drawing that was immediately followed by a mask (150 or 500 ms). Participants were instructed to press one mouse button if they thought the image was whole and another if they thought it was scrambled (counterbalanced across participants).

The whole stimuli were black and white line drawings of common objects. Each image had an 8×8 black grid superimposed onto it to make the task of distinguishing it from a scrambled image more difficult. Unbeknownst to the participants, the experiment was divided into 2 blocks of 256 trials. During the second block, many of the same stimuli were presented again either in the same position or in one of 2 equidistant positions either within the same hemifield or in the opposite hemifield (Fig. 1). With the short duration (66 ms) stimuli, no participants reported any awareness of the repeats when questioned.

Retinotopically matched scrambled images were generated from these whole images via the following method. Each whole image was cut into 64 equal sized squares (along the grid lines) and each square marked as blank or containing a line. The retinotopically matched scrambled image was generated by iterating through each square that contained a line and filling it with a corresponding square from a randomly selected unscrambled image. This process generated a scrambled image that contained lines from a number of whole images, all in their appropriate positions, whose retinotopic envelope roughly matched the original whole image. This process was then repeated over the entire set of whole images without replacement, ensuring that no line was repeated over the set of scrambled images. Thus, the scrambled set contained all the same line segments as the unscrambled set, in the same positions but scrambled across images.

Masks were generated by taking a random set of 16 scrambled images, inverting them, and then overlaying them onto one another. This process was repeated for each trial such that no mask was ever repeated during the experiment, eliminating the possibility of any learning of the mask. Selection was balanced to ensure that each scrambled image was included an equal number of times over the entire set of masks. These masks were extremely effective such that only 3 participants could report the identity of any whole image with the short duration (66 ms) stimuli.

A subset of objects was designated to serve as controls for each participant. The control objects were divided in half and presented during the first and second blocks to provide a measure of general improvement on the task. The set of control objects was counterbalanced across participants such that every object served an equal number of times as a first and second block control object. Those objects not used as control trials were then randomly assigned to one of the 3 other conditions (Within-Position, Within-Field, Between-Fields).

As the repeat presentations were separated from the initial presentations by an average of 256 trials, it is unlikely that an attentional, semantic, or low-level confound is affecting our results. Each trial contains an object (whole or scrambled) and a mask, consisting of many of the same low-level features (e.g., line segments, curves) found in every stimulus in the experiment. Any stimulus-specific low-level priming would be interfered with heavily by these intervening trials, 64 of which occur in the same position as any particular initial presentation. Thus, any low-level priming should be nonspecific and well captured by the control trials. Attentional confounds are also unlikely as stimuli occurred in each position with equal frequency, making any spatial prioritization untenable. Attention could, over time, be better focused on the 4 positions where stimuli did occur but that would lead only to nonspecific improvements in performance. Finally, any explicit semantic strategy (e.g., remembering the names of whole objects) is also unlikely as the list to be remembered would be 128 object names long.

Significance of priming effects and decrements in priming were established with one-tailed tests as every test was planned and had a clear hypothesis associated with it.

Event-Related fMRI Stimuli and Task

During the event-related runs of the fMRI experiment, participants were presented with a subset of the whole images used in the behavioral experiments. Four new face line drawings were also presented (Supplementary Item 2). These images were presented in at each of the retinotopic positions used in the behavioral experiment. Each image was presented for 300 ms. The use of line drawings, rather than full color images, left only shape information available in the stimuli (Fig. 2). This, combined with the very weak categories present in our stimuli, makes it harder to find categorical effects but more likely that we will be able to tell individual stimuli from one another (individuation). Individuation is not often measured, due to the fact that measuring it requires experimental designs that do not assume category structure, as has been assumed in previous block-design studies (Sayres and Grill-Spector 2008; Schwarzlose et al. 2008; Williams et al. 2008). Our design makes no assumptions about category structure or position sensitivity, allowing us to measure the contributions of each to the response of object-selective cortex in an unbiased way. Furthermore, our design provides the most direct and conservative test of the position-dependence of object information. Previous studies have tested categorization and found no effect of position changes (Sayres and Grill-Spector 2008; Schwarzlose et al. 2008; Williams et al. 2008), but this null finding might be attributable to the relative ease of categorization as compared with individuation as a measure of object information (see Supplementary Item 1 for further discussion).

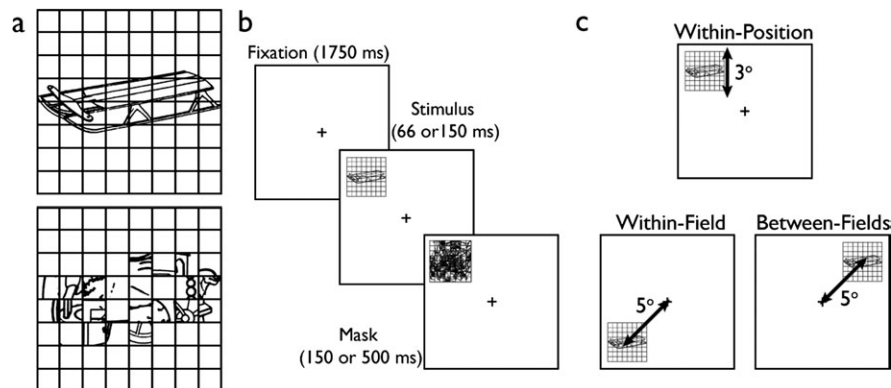


Figure 1. Behavioral priming paradigm. (a) Example line drawing of an object and its retinotopically matched scrambled image. The scrambled images had retinotopic envelopes that matched the whole objects as closely as possible, and gridlines were superimposed on all images to prevent the use of low-level cues (see Materials and Methods for information on image scrambling). (b) Trial presentation sequence and timing. After a 150-ms fixation period an image was presented in one of the 4 quadrants of the visual field for either 66 or 150 ms followed by a trial unique mask (150 or 500 ms, respectively). Two blocks of trials, each containing 256 trials were presented. (c) Illustration of the 3 possible priming conditions in the behavioral experiment, given a presentation in the first block at the location depicted in (b). In the second block of trials, an image could be repeated in the same position as in the first block (within-position), in a different position in the same visual field (within-field) or in the opposite visual field (between-fields). The numbers reflect stimulus size and distance from the fovea in degrees of visual angle.

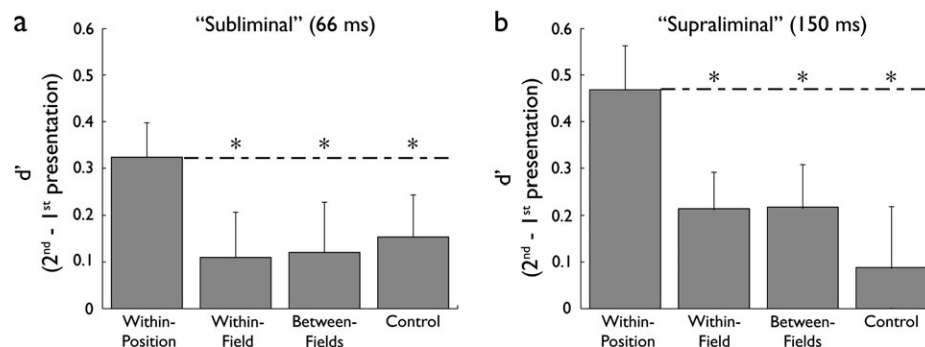


Figure 2. Behavioral priming results. Results are shown separately for the 2 stimulus timings. Data are presented as the difference in d' values between the first and the second block of trials. A significant difference from Control indicates significant priming above and beyond any nonspecific task improvement. (a) Sixty-six milliseconds presentations. Significant priming was observed only within-position. Furthermore, there was significantly greater priming within-position than within-field or between-fields indicating position-dependence of priming. (b) One hundred and fifty milliseconds presentations. As with the shorter presentations, significant priming was observed only within-position. Furthermore, there was significantly greater priming within-position than between-fields. Although the difference between within-position and within-field was not significant at this timing, perhaps indicating greater transfer of priming, the interaction between condition and stimulus timing was not significant ($P > 0.1$). Error bars show the between-subjects standard error. * Indicates a significant reduction ($P < 0.05$) in priming from the within-position condition.

In order to encourage participants to attend to the stimuli and to maintain fixation, participants performed a color-matching task between the fixation cross and the peripheral stimuli. The fixation cross changed from white to one of 4 possible colors at the same time as the stimulus appeared. Participants reported whether or not the color of the fixation cross and the stimulus matched. All colors were counterbalanced such that they occurred equally often in each of the 4 positions. We used this task, which was orthogonal to both the position and identity of the objects, to reduce the possibility of any task confounds.

fMRI Localizer Stimuli and Task

Three independent scans were also collected in each participant to localize lateral occipital (LO), posterior fusiforms (PFs), and fusiform face area (FFA). Each of these scans was an on/off design with alternating blocks of stimuli presented while participants performed a one-back task. LO and PFs were localized using the contrast of objects minus scrambled objects and FFA with the contrast of faces minus objects. Object and face images were grayscale photographs. Scrambled objects were generated via the same method as the scrambled images in the behavioral experiment, with the exception that the images were cut into 400 rather than 64 squares. This method of scrambling produced retinotopically matched stimuli to compare with the object images, reducing the chance that purely retinotopic voxels would be included in the LO and PFs regions of interest (ROIs).

fMRI Scanning Parameters

Participants were scanned on a research dedicated GE 3-Tesla Signa scanner located in the Clinical Research Center on the National Institutes of Health campus in Bethesda. Partial volumes of the temporal and occipital cortices were acquired using an 8-channel head coil (22 slices, $2 \times 2 \times 2$ mm, 0.02 mm interslice gap, time repetition [TR] = 2 s, time echo = 30 ms, matrix size = 96×96 , field of view = 192 mm). In all scans, slices were oriented approximately perpendicular to the calcarine sulcus. Six event-related runs (263 TRs), 6 localizer scans (80 TRs), and high-resolution anatomical images were acquired in each session.

fMRI Statistical Analysis

FFA, LO, and PFs ROIs were generated in each hemisphere for each participant from the localizer runs. Significance maps of the brain were computed by performing a correlation analysis thresholded at 0.0001 (uncorrected). ROIs were generated from these maps by taking the contiguous clusters of voxels that exceeded threshold and occupied the appropriate anatomical location based on previous studies (Sayres and Grill-Spector 2008; Schwarzlose et al. 2008).

Significance maps in the event-related runs were created by performing t -tests between each condition and baseline. The t -values for each condition were then extracted from the voxels within each ROI and cross correlated (Haxby et al. 2001; Chan et al. 2010). We used t -values rather than coefficients as they tend to be slightly more stable, reducing the impact of noisy voxels that may nonetheless have large coefficients associated with them. Our results remain the same when coefficients were used. This yielded matrices that represent the similarity in the spatial pattern of response across the ROI between each pair of conditions.

To establish whether position effects were significant, 2 types of tests were performed. In the standard analysis, the matrices were averaged by position and these values compared via t -tests. Significance of imaging effects were established with one-tailed tests as every test was planned and had a clear hypothesis associated with it. In the permutation analysis, a random proportion of the similarities from the unaveraged matrices were switched between the 2 conditions being compared and the averages by position then calculated. The randomization reflects the null hypothesis that the 2 were equivalent, implying that the data generated by those conditions is interchangeable. This procedure was repeated 10 000 times to derive the distribution of differences between conditions that might have arisen by chance fluctuations between 2 identical conditions. If the observed difference was greater than 95% of the random distribution, then there was less than a 5% chance of the observed difference arising randomly, and the difference was determined to be significant.

All of our correlation values were well below 0.5 and above -0.5, making it unlikely that their distributions were not normal. Nonetheless, we reanalyzed our data with Fischer transforms and found no impact on either our position or individuation effects.

Results

Object Priming Reduces with Changes in Position

We used behavioral priming to investigate the position-dependence of visual object representations (Fig. 1). Priming is defined as the improvement in performance during the second presentation of a stimulus compared with the first. Priming is thought to reflect the potentiation of the stimulus representation by the first presentation and is widely used to measure the degree of overlap between representations for words (e.g., monkey-tree vs. monkey-sedan); here, it is used to measure the overlap between representations of an object across positions.

Previous priming studies investigating position-dependence (Biederman and Cooper 1991; Bar and Biederman 1998, 1999;

Fiser and Biederman 2001) have required participants explicitly name each stimulus. These naming tasks inherently engage semantic processing making it difficult to assess purely “visual” priming (see Discussion). To avoid the possibility of a semantic confound in the present study, participants made a nonverbal discrimination (button press), indicating whether a briefly presented stimulus was a whole or scrambled object (Fig. 1*a*).

Unbeknownst to the participants, trials were divided into 2 sequential blocks each containing one presentation of each stimulus (128 whole and 128 matched scrambled stimuli). The second presentation of any given stimulus occurred after an average of 256 trials (range 224–288). The large number of intervening trials dramatically reduces the possibility of low-level, attentional, or semantic confounds being responsible for any observed effects (see Materials and Methods). Within each block, there were 64 control stimuli that were only seen once, providing a measure of any nonspecific improvements in performance (e.g., task learning) between blocks.

Previous priming studies have reported no reduction in priming with changes in position (Biederman and Cooper 1991; Fiser and Biederman 2001) with 150 ms (supraliminal) presentations and have been used to argue for position-independent object representations. However, shorter durations (66 ms, “subliminal”) have revealed reduced priming with position shifts (Bar and Biederman 1998, 1999).

Here, we ran our nonverbal discrimination with both subliminal (66 ms) and “supraliminal” (150 ms) timing on 2 groups of participants ($n = 27$: subliminal; $n = 31$: supraliminal). Outlier participants were determined using 2 criteria. First, d' scores which were more than 2 standard deviations from the population mean ($n = 1$: supraliminal). Second, differences in d' between the first and second presentations that were greater than 2 standard deviations from the mean in any condition ($n = 1$: supraliminal).

For both stimulus durations, discrimination performance improved during the second presentation (66 ms: first $d' = 0.55$, second $d' = 0.73$; 150 ms: first $d' = 1.55$, second $d' = 1.88$ see Supplementary Item 3 for raw scores). To assess priming, we input the difference scores between the first and second presentations (Fig. 3*a,b*) into an omnibus analysis of variance (ANOVA) with Condition (Within-Position, Within-Field, Between-Fields, and Control) as a within-subject factor

and Experiment (66, 150 ms) as a between-subjects factor. There was a significant main effect of Condition ($F_{3,156} = 3.696$, $P < 0.05$) but no main effect ($P > 0.3$) or interactions ($P > 0.6$) involving Experiment, indicating that stimulus duration had no significant effect on the observed pattern of results.

Significant priming was defined as greater improvement in performance on repeat trials than on control trials. A series of planned comparisons revealed significant priming ($t_{1,53} = 2.873$, $P < 0.01$, one-tailed) only when the first and second presentations occurred in the same position (within-position), with no significant priming observed when position changed (either within- or between-fields) ($P > 0.5$). Furthermore, there was significantly greater priming within-position than with position shifts either within field ($t = 2.949$, $P < 0.01$, one-tailed) or between-fields ($t = 2.975$, $P < 0.01$, one-tailed). There was no significant difference in priming for within-field versus between-fields position shifts ($t = 0.075$, $P > 0.9$).

Thus, when an object is visually primed, there is at best limited transfer of priming to different positions (see also Kravitz et al. 2008). Our results demonstrate that the behavioral output of high-level object representations are position-dependent, and we now turn to fMRI to directly confirm the position-dependence of visual object representations.

How do Position and Object Identity Affect the Response of Object-Selective Cortex?

We used an iterative variant (see Supplementary Materials) of split-half correlation analysis (Haxby et al. 2001) to investigate how changes in position affect the spatial pattern of response across object-selective cortex. In an event-related fMRI paradigm, 10 participants saw line drawings of 24 objects (Supplementary Item 2) in each of 4 positions used in the behavioral experiment (96 total conditions), while performing an orthogonal color-matching task (Fig. 2) between the fixation cross and stimuli (93% accuracy, see Materials and Methods). Two object-selective ROIs were defined in each hemisphere of each participant using independent localizer data (LO, PFs; see Supplementary Item 4). The split-half analysis of the event-related data yielded a 96×96 similarity matrix for each ROI wherein each point represents the correlation or similarity between a pair of conditions (Kriegeskorte et al. 2008; Drucker and Aguirre 2009). The structure of the similarity matrix allows

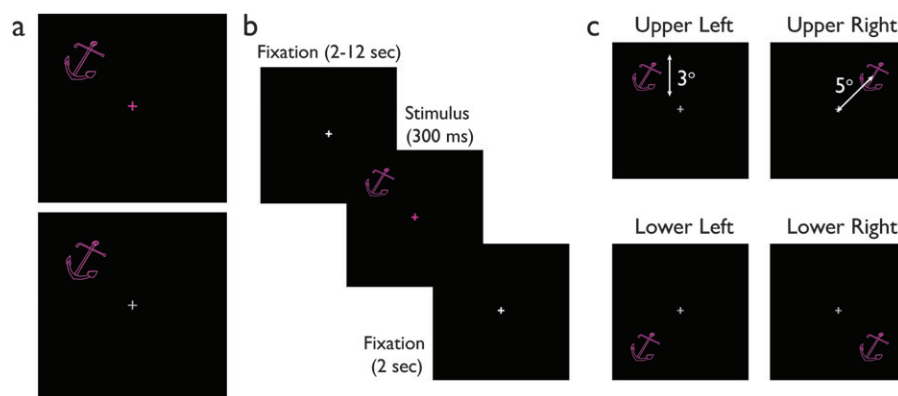


Figure 3. Imaging paradigm. (*a*) Participants were asked to indicate whether the color of the lines in the object and the color of the fixation cross were the same or not using a button-box. This task encouraged the participants to maintain fixation while simultaneously attending to the peripheral stimulus. The task was also orthogonal to both the position and identity of the object images. Note that stimuli and fixation cross could be any one of 4 colors (green, blue, gray, and purple), not only the one presented. Colors were counterbalanced across positions and match/nonmatch. (*b*) Stimuli were presented for 300 ms with a variable interstimulus interval from 3.7 to 12 s. (*c*) Stimuli occurred in each of the 4 positions used in the behavioral experiment.

us to assess the different contributions of identity and position to the pattern of response in object-selective cortex (Supplementary Item 5).

Here, we first assess the effect of position changes on the patterns of response in LO and PFs before considering whether the position effects constrain object identity information.

Object-Selective Cortex Is Extremely Sensitive to Changes in Position

The full similarity matrix for an example ROI (Left PFs, Fig. 4a) shows position to be the primary determinant of correlation across conditions (see Supplementary Item 6 for other ROIs). Averaging the full matrix by position (Fig. 4b) to produce a 4×4 position matrix reveals much higher correlations within-position than between-positions, confirming strong sensitivity to position.

All 4 ROIs show higher correlations within- than between-positions (Fig. 5a), indicating strong effects of position changes (see Supplementary Item 7 for FFA). Each ROI showed the same ordering of similarity (Within-Position, Within-Field, Between-Fields, and Opposite) (Fig. 5b, see Supplementary Item 8 for schematic explanation of Fig. 5), which matched the reduction in priming observed in the behavioral experiments (Fig. 3). An omnibus ANOVA with ROI (left PFs, right PFs, left LO, and right LO) and Position (Within-Position, Within-Field, Between-Fields, and Opposite) as repeated measures revealed a highly significant effect of Position ($F_{3,27} = 21.352$, $P < 0.001$) and no interactions or main effects involving ROI. Pairwise comparisons and permutation tests (see Materials and Methods) revealed significant ($P < 0.01$) reductions in correlation for all position shifts compared with within-position. Although some prior studies have reported partial overlap of posterior object-selective cortex (LO) with retinotopic cortex (Sayres and Grill-Spector 2008; Arcaro et al. 2009), none of our ROIs included any significant proportion of retinotopic voxels (Supplementary Item 9).

Position Sensitivity is Greater in the Contralateral than Ipsilateral Field

Consistent with prior results from both macaque and human physiology (Niemeier et al. 2005; Hemond et al. 2007), we found stronger ROI-average activation to contralateral than ipsilateral stimuli (Fig. 5c). An omnibus ANOVA with ROI (left PFs, right PFs, left LO, right LO) and Hemifield (Ipsilateral,

Contralateral) as within-subject factors revealed a significant main effect of Hemifield ($F_{1,9} = 7.221$, $P < 0.05$) and a main effect of ROI ($F_{3,27} = 3.909$, $P < 0.05$), reflecting a greater response in PFs than in LO. No interaction between Hemifield and ROI was observed ($P > 0.15$).

Like the overall activation, within-position correlations were also greater in the contralateral than ipsilateral hemifield, but there was an asymmetry between the upper and lower quadrants (see analysis directly below).

To analyze the effect of laterality on the spatial pattern of response in each ROI, we compared the within-field correlations in the contralateral and ipsilateral field separately (Fig. 5d). In each ROI (except right LO which showed the same trend), there were significantly greater ($P < 0.05$ Bonferroni corrected; permutation test) correlations within the ipsilateral than contralateral hemifield, suggesting less sensitivity to position in the ipsilateral field. This finding is consistent with prior studies of macaque IT cortex showing that neurons responsive to ipsilateral stimuli generally have very large RFs (Op De Beeck and Vogels 2000), which tend to encompass both the upper and lower quadrants.

Thus, responses are stronger and more sensitive to position in the contralateral than ipsilateral visual field.

How do the Representations of the Quadrants Differ across the ROIs?

For each ROI, there was one quadrant, always in the contralateral visual field, where stimulus presentations produced the greatest within-position correlation (Fig. 5a, blue asterisks; Fig. 5e, red boxes). Permutation tests confirmed that in each ROI, save one quadrant in right PFs, the preferred quadrant produced significantly greater within-position correlations than all other quadrants ($P < 0.05$, Bonferroni corrected). Note that in PFs the strongest correlations were in the upper contralateral quadrant, whereas in LO, the strongest correlations were found in the lower contralateral quadrant. Note also that the strongest activation was also observed in same quadrants for each ROI (red boxes; Fig. 5d), though these differences were not significant.

A lower field bias has been previously reported in LO (Sayres and Grill-Spector 2008; Schwarzlose et al. 2008), but to our knowledge, this is the first demonstration of an apparent upper field bias in PFs. It is possible that these biases result from the

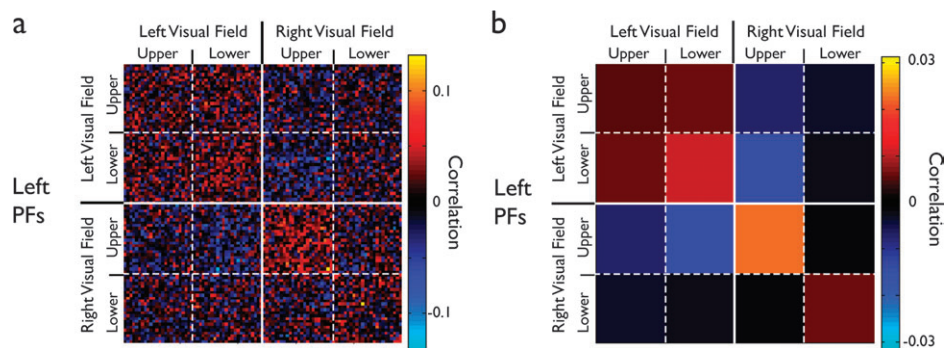


Figure 4. Example similarity and position matrix. (a) Raw similarity matrix for the left PFs ROI averaged across all 10 participants. The matrix is 96×96 (24 objects \times 4 positions), with each point reflecting the amount of correlation between a pair of conditions across the 2 halves of the data. The main diagonal from the upper left to lower right corner are the correlations between a condition and itself in the 2 halves of the data. The matrix is ordered by position as indicated on the axes. Solid lines denote borders between hemifields. Dashed lines denote borders between upper and lower quadrants. Note that the positive and negative correlations are well grouped by this position ordering, suggesting position is the primary determinant of the pattern of response. (b) The matrix in (a) averaged by position.

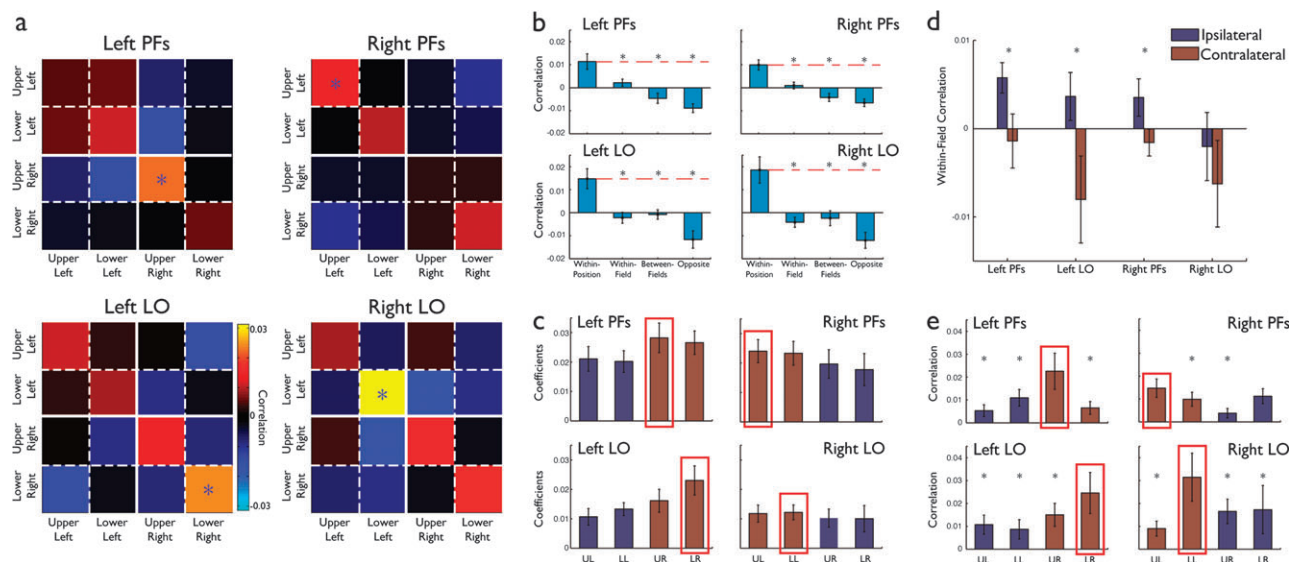


Figure 5. Imaging results: position effects. (a) Average matrices as in Figure 3b for all 4 object-selective ROIs. The blue asterisks indicate the highest correlation present in each matrix. Summary data from this panel is replotted in (b), (d), and (e). The color scale of each matrix is identical as indicated by the bar next to the bottom left plot. (b) The average correlation in each ROI between stimuli within-position and between different positions (rows in the matrices in [a]). Each possible position shift from the behavioral experiment is present in the imaging results, with the addition of the opposite position, which is a shift both between hemifields and upper and lower quadrants. For example, the within-position correlation is the average of the 4 diagonal values in (a). Red lines indicate the expected level of correlation if position has no impact on the spatial pattern of response in each ROI. * Indicates a significant reduction ($P < 0.05$) in correlation from this expected performance as revealed by a permutation test (see methods). (c) Plots of the average coefficients in each ROI for stimuli at each position. Red boxes indicate the highest response, which was always in the contralateral field and in the lower visual field for LO and the upper visual field for PFs. (d) The average correlation for within-field position shifts (correlation between the upper and lower quadrants) in each ROI as a function of whether that shift is in ipsi- or contralateral space. * Indicates a significant difference between the ipsi- and contralateral correlations within the ROI. The higher correlation values for ipsi- compared with contralateral space indicate less sensitivity to changes in position in ipsilateral space. (e) Plot of the within-position correlations (diagonal values in [a]) across the ROIs. The red boxes indicate the preferred position of the ROI (highest correlation) and correspond to the blue asterisks in (a). As for the magnitude of response (c), the highest correlations are always in the contralateral field and in the lower visual field for LO and the upper visual field for PFs. * Indicates a significant difference between the correlation at a position and the correlation at the preferred position. Error bars in each panel reflect the between-subjects standard error.

proximity of these regions to the early visual cortex representations of the upper and lower visual fields. The upper field representations are inferior to the calcarine sulcus and lie closer to PFs, while the lower field representations are superior to the calcarine sulcus and lie closer to LO.

Thus, each of the ROIs we examined shows the strongest correlations in a separate quadrant of the visual field with a preference for the lower field in LO and for the upper field in PFs.

Thus, both LO and PFs are highly sensitive to shifts in position both within- and between-hemifields. We next demonstrate that this position information interacts with object identity information, leading to significantly reduced object individuation across positions.

Object Individuation Is Significantly Reduced with Changes in Positions

To quantify individuation, we compared the within-object and between-object correlations (Haxby et al. 2001) both within-position and between-positions. If the within-object correlation is significantly greater than the average of the between-object correlations in the same position, then that object can be individuated within-position. Between-position individuation is defined as the difference between the within-object correlations and the average of the between-object correlations across positions (Fig. 6a), similar to previous studies (Schwarzlose et al. 2005; Sayres and Grill-Spector 2008).

To establish the degree of within-position individuation, within- and between-object correlations were calculated for

each of the 24 objects in the contra- and ipsilateral hemifields and collapsed across hemisphere (Fig. 6b-e: left panels). Between-position individuation for each of the 24 objects was calculated by averaging across the 3 possible between-position comparisons (Fig. 6b-e: right panels) (unaveraged data for each position change and ROI are shown in Supplementary Item 10). Average individuation performance across the set of 24 objects (Fig. 6f,g) were entered into an omnibus ANOVA with Region (PFs, LO), Hemifield (Ipsilateral, Contralateral), and Position (Within, Between) as within-subject factors. Note that the null hypothesis here is that of complete position invariance, no difference in individuation with changes in position. This test revealed a significant main effect of Position ($F_{1,10} = 5.139$, $P < 0.05$), indicating a significantly reduced ability to individuate objects with changes in position. Furthermore, planned comparisons revealed significant individuation only within-position in the contralateral hemisphere for both PFs ($P < 0.001$) and LO ($P < 0.05$). No other combination of Hemifield and Position showed significant individuation ($P > 0.18$). This result shows that these regions do not code object identity and position independently, rather object representations are tied to particular positions. Our individuation results do not simply reflect categorization (Supplementary Item 11, 12), which is unsurprising given the heterogeneous stimuli that comprise the categories in our stimulus set (Supplementary Item 2). Importantly, the reduction in individuation across positions is not a negative result, but a significant difference between 2 conditions that are completely equivalent in every way except for whether individuation is occurring within- or

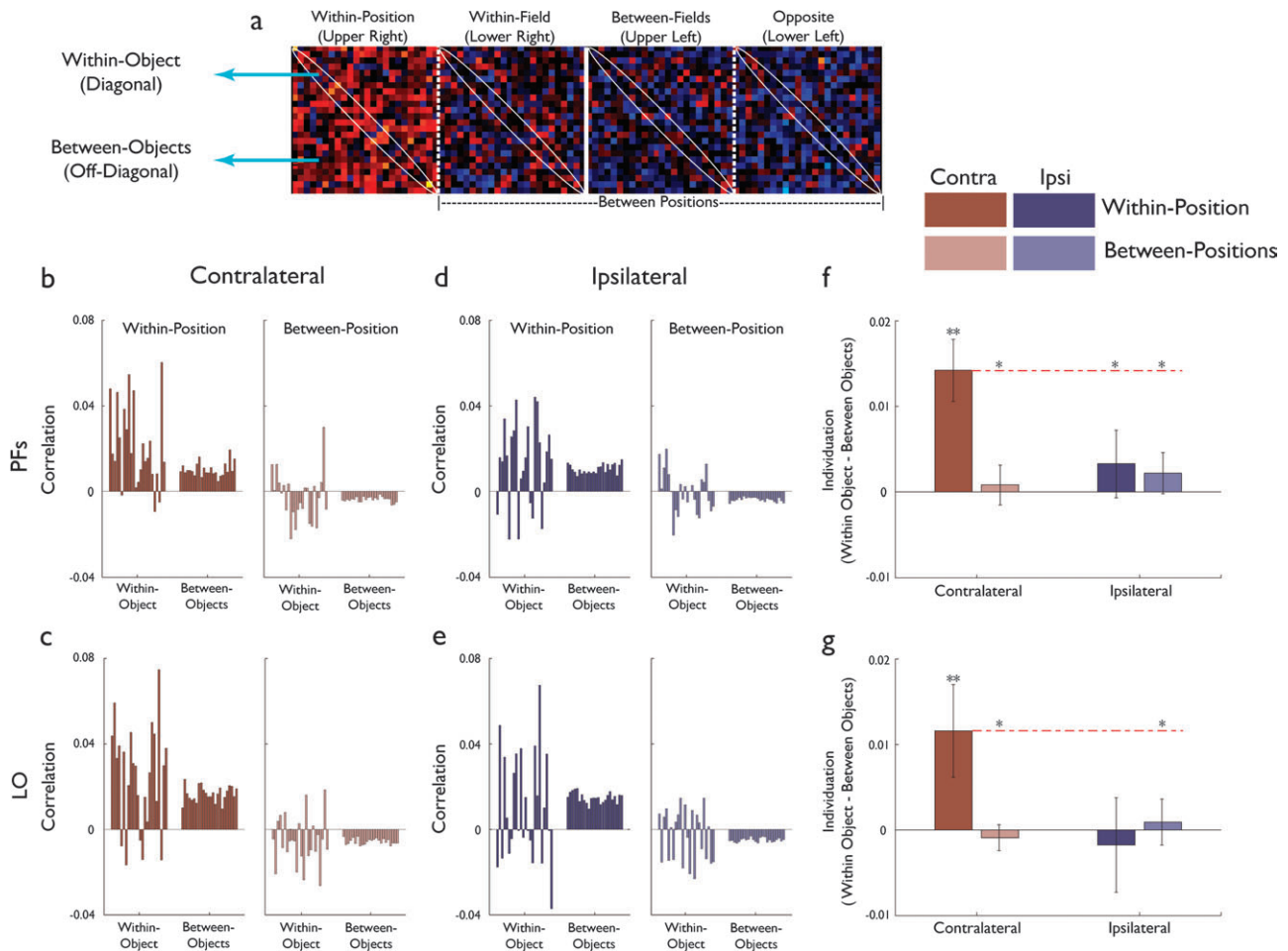


Figure 6. Imaging results: individuation effects. Schematic of the individuation analysis logic. The data are taken from the third row of the response of left PFs (Fig. 4a) and is color scaled the same way. Individuation within a position is defined as the average correlation between each object and itself (within-object) (far left box; white ellipse) against the average of correlation between each object and the other objects presented in the same position (average across the rest of the row within the box [between-objects]). Greater within- than between-objects correlations indicate individuation. Individuation across positions is established through the same calculation but position is no longer constant. Within-object correlation across positions is defined as the average correlation between each object in one position (e.g., upper right as shown) and the same object in a different position (right boxes; white ellipses). Between-object correlation across positions is the average correlation between each object in one position and the other objects in the other position. (b) Within- and between-objects correlations averaged across left and right PFs and the upper and lower quadrants of the contralateral hemifield. (c) Same as (b) but for the average of left and right LO. (d) Same as (b) but now averaged from the upper and lower quadrants of the ipsilateral hemifield. (e) Same as (d) but for the average of left and right LO. (f) Difference scores resulting from the subtraction of the within- and between-objects correlation shown in (b, d). In this plot, a significant difference from zero (**) reflects significant individuation ($P < 0.05$). The dashed red line represents predicted performance if changes in position (lighter bars) for a given hemifield (contralateral, ipsilateral) have no impact on individuation. * Indicates a significant reduction in the ability to individuate ($P < 0.05$). (g) Same as (f) but for LO. Error bars represent the between-subjects standard error in (f, g).

between-positions. We do not claim that there is no individuation for objects across positions, though we find no evidence of it. Rather, this result shows that, contrary to the predictions of invariance, there is “significantly” reduced individuation across positions.

Furthermore, the omnibus ANOVA also revealed a highly significant interaction between Hemifield and Position ($F_{1,10} = 10.583$, $P < 0.01$), reflecting a reduction in individuation in the ipsilateral field. To directly test for this effect, within-position individuation scores were entered into an ANOVA with Region (PFs, LO) and Hemifield (Ipsilateral, Contralateral) as within-subject factors, revealing a main effect of Hemifield ($F_{1,10} = 5.44$, $P < 0.05$). Further pairwise comparisons revealed reductions in individuation in the ipsilateral hemifield in both PFs ($P < 0.05$) and LO ($P = 0.07$).

Thus, object individuation was significantly stronger within-position than between-positions and in the contra-

ipsilateral field. This pattern of results indicates an interaction between position and object information and suggests position-dependent visual object representations. The results of this object individuation analysis are entirely consistent with the results of the behavioral priming and suggest that even high-level visual object representations are tied to limited portions of the visual field.

Discussion

We have provided converging evidence with both behavior and fMRI that visual object representations are position-dependent. Behaviorally, visual object priming was significant only when position was unchanged, with position shifts leading to decreased priming. Direct measurement of the object representations in object-selective cortex demonstrated the position-dependence of object representations with both 1)

large changes in the pattern of response following position shifts both within- and between-hemifields and 2) significantly reduced object individuation with changes in position (see Fig. 7 for a graphic summary). This convergence between the imaging and behavioral results is critical (see also Supplementary Item 13). The imaging results demonstrate that the behavioral effects reflect the representations in object-selective cortex. The behavioral results suggest that our multivariate measures reflect relevant aspects of the neural response. Overall, our results suggest that object representations are tied to limited portions of the visual field. The precise extent of this position-dependence remains an open question but at least for the ~ 7 degree changes in position we tested, there is a significant difference in the visual representations. Given the distribution of RFs observed in IT (Op De Beeck and Vogels 2000), it is likely that the strength of object information (e.g., priming and individuation) will be graded, decreasing with increasing distances.

Behavioral Position-Dependence

Our finding of reduced priming with shifts in position stands in contrast to one widely cited study reporting position-independent priming between hemifields in an object-naming task (Biederman and Cooper 1991) (see also Fiser and Biederman 2001). However, it is possible that the use of an explicit naming task may have led to an overestimation of the degree of position-independence due to the engagement of semantic representations. The authors argued against an influence of semantics by showing a reduction in priming to new objects with the same category names (semantic controls), suggesting that some proportion of the effects were visual and not semantic. However, the use of the semantic controls makes the assumption that visual and semantic representations are independent and additive, which is not

necessarily the case. For example, engagement of semantic representations may actually reduce the effects of visual priming, possibly through feedback. Ultimately, achieving a pure measure of visual priming requires a paradigm that engages semantics as little as possible. The results from our nonverbal discrimination and fMRI strongly suggest that the observed position effects emerge from visual object representations. Furthermore, our findings are consistent with other behavioral studies that have investigated the effects of position changes on visual representations (Nazir and O'Regan 1990; Dill and Fahle 1998; Dill and Edelman 2001; Afraz and Cavanagh 2008) (see also Kravitz et al. 2008).

Position-Dependence in Object-Selective Cortex

Our finding of position-dependent object representations in object-selective cortex stands in apparent contrast to prior studies in both human and nonhuman primates (Hung et al. 2005; Sayres and Grill-Spector 2008; Schwarzlose et al. 2008; Williams et al. 2008). Two factors may explain the lack of evidence for position-dependent object representations in these studies. First, most these studies (Williams et al. 2007; Schwarzlose et al. 2008) tested position shifts only within a hemifield, where the effect of position is weaker, and not between hemifields. For example, categorization and individuation of objects have been reported (Hung et al. 2005) across shifts of position of up to 4 degrees within-hemifield using the population response across a large set of macaque IT neurons. It is worth noting, however, that there was some decrement in performance with even such small position changes. Although another study (Sayres and Grill-Spector 2008), did include position shifts both within- and between-hemifields, most position shifts were within-hemifield, and data were collapsed across both types of shift when determining the degree of position-dependence. Second, all the fMRI studies have used

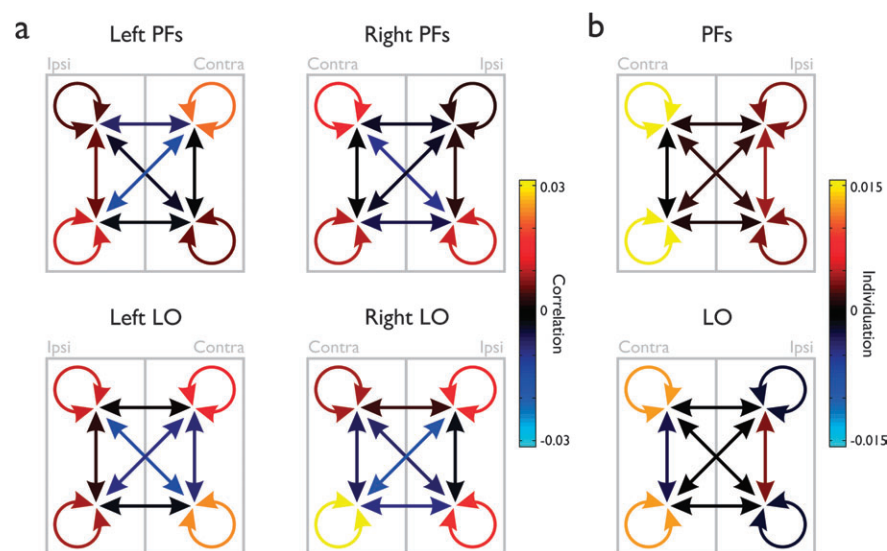


Figure 7. Imaging results: position and individuation effects summary. This figure places the pattern of correlations observed across the object-selective ROIs (Fig. 4a) back in context with the positions of the stimuli. Each straight arrow represents the correlation or individuation between-positions. Each curved arrow represents the correlation or individuation within-position. The color of the arrows reflects the strength of the correlation or individuation (colorbar on far right). (a) Position effects. Note that the within-position correlations are generally much stronger than the between-position correlations, especially when the 2 positions are in different hemifields (Fig. 4b). Note also the preferred positions of each ROI lies in one quadrant of the contralateral visual field (PFs: upper; LO: lower) such that over all the ROIs all 4 object positions are covered (Fig. 4d,e). (b) Individuation effects. Note that data were averaged over the upper and lower quadrants as in Figure 5 as there was not enough power to establish effects within each individual quadrant. So the upper and lower portions of each panel are identical. The only condition for which there is significant individuation is within-position in the contralateral hemifield for both LO and PFs.

block-design experiments and have focused on categorization among a small number of categories rather than individuation (Sayres and Grill-Spector 2008; Schwarzlose et al. 2008; Williams et al. 2008), both of which might lead to a failure to find position-dependence (Supplementary Item 1).

It is possible that our population measure is not sensitive to those few neurons which have larger RFs (Op De Beeck and Vogels 2000), but ultimately, the behavioral results in the current study suggest even if there is some ability to read out independent of position based on these neurons, the output of these visual object representations is position-dependent. From our results, it appears that visual object representations in human object-selective cortex closely match the average RF properties of neurons observed in macaque IT (which may also constrain position-independence at the population level [Goris and Op de Beeck 2009]). From this correspondence, it is likely that visual object representations span larger portions of the visual field than do early visual representations but still maintain some relative degree of position-dependence.

The Role of Experience in Position-Dependence

Our study included no systematic manipulation of experience with the objects. Our stimuli were drawn from a set of highly familiar objects, suggesting that even highly familiar visual representations are position-dependent. However, experience is likely to play an important role in the degree of position-dependence of visual representations. In particular, a series of behavioral and neurophysiology experiments have shown that position-dependence can be manipulated either by providing visual experience in one location only (Cox and DiCarlo 2008) or by disrupting the connection between the foveal image of an object and the peripheral image (Cox et al. 2005; Li and DiCarlo 2008). It may be that increased position-independence takes the form of learned associations between position-dependent representations of the same object in different positions (Miyashita 1988; Wallis and Rolls 1997; Cox et al. 2005; Li and DiCarlo 2008).

In our studies, we investigated position-dependence at 4 equally eccentric peripheral locations. However, the fovea may be a critical retinal location for integrating object representations across position. Given that objects are typically foveated, visual experience may principally produce associations between peripheral and foveal images. Furthermore, the fact that object recognition principally occurs at the fovea may obviate the need for position-independent representations across peripheral portions of visual space and provide a behavioral compensation for position-dependence.

Implications

These results have implications for the interpretation of perceptual learning effects. Typically these effects are assumed to arise in low-level visual cortex if they evidence any position-dependence (Golcu and Gilbert 2009). If even high-level object representations evidence some position-dependence, great care is necessary when ascribing neural substrates to behavioral effects based solely on the degree of position-dependence. Future work will need to establish more precisely the exact degree of position-dependence at each level of the visual hierarchy.

Position-dependent high-level object representations also imply that lateralized damage to object-selective cortex should

selectively impair peripheral object processing primarily in only one visual field. Visual field has an enormous impact on both priming and individuation, suggesting that, as observed in monkey IT, 2 largely distinct populations of neurons process peripheral objects in the 2 fields. This structure suggests that lateralized damage should lead to position-specific object agnosia, particularly when every effort is made to measure only visual object processing.

Conclusion

Our results demonstrate that retinotopic position strongly modulates visual representations in high-level object-selective cortex. These results suggest that there are at least partially nonoverlapping representations of the same object in different positions. This contradicts the widely held assumption that visual object representations are position-independent, an assumption that underlies many models of visual object recognition. The pervasiveness of position even in high-level visual object representations is perhaps less surprising when we consider that spatial structure is a core constituent of all visual experience.

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

Funding

National Institute of Mental Health Intramural Research Program.

Notes

Thanks to Marlene Behrmann, Nancy Kanwisher, Alex Martin, Hans Op de Beeck, Rebecca Schwarzlose, Sandra Truong, Leslie Ungerleider, Mark Williams, and members of the Laboratory of Brain and Cognition, National Institute of Mental Health for helpful comments and discussion. Dwight Kravitz designed, implemented, and ran the studies, analyzed the data, and wrote the manuscript. Nikolaus Kriegeskorte helped in designing the imaging experiments and analyses. Chris Baker supervised the entire set of studies and helped to write the manuscript.

Conflict of Interest: None declared.

References

- Afraz SR, Cavanagh P. 2008. Retinotopy of the face aftereffect. *Vision Res.* 48:42–54.
- Arcaro MJ, McMains SA, Singer BD, Kastner S. 2009. Retinotopic organization of human ventral visual cortex. *J Neurosci.* 29: 10638–10652.
- Bar M, Biederman I. 1998. Subliminal visual priming. *Psychol Sci.* 9:464–469.
- Bar M, Biederman I. 1999. Localizing the cortical region mediating visual awareness of object identity. *Proc Natl Acad Sci U S A.* 96: 1790–1793.
- Biederman I, Cooper EE. 1991. Evidence for complete translational and reflectional invariance in visual object priming. *Perception.* 20:585–593.
- Chan AW, Kravitz DJ, Truong S, Arizpe J, Baker CI. 2010. Cortical representations of bodies and faces are strongest in their commonly experienced configurations. *Nat Neurosci.* Advance Access published March 7, doi: 10.1038/nn.2502.
- Cox DD, DiCarlo JJ. 2008. Does learned shape selectivity in inferior temporal cortex automatically generalize across retinal position? *J Neurosci.* 28:10045–10055.

- Cox DD, Meier P, Oertelt N, DiCarlo JJ. 2005. 'Breaking' position-invariant object recognition. *Nat Neurosci*. 8:1145-1147.
- DiCarlo JJ, Cox DD. 2007. Untangling invariant object recognition. *Trends Cogn Sci*. 11(8):333-341.
- DiCarlo JJ, Maunsell JH. 2003. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol*. 89:3264-3278.
- Dill M, Edelman S. 2001. Imperfect invariance to object translation in the discrimination of complex shapes. *Perception*. 30:707-724.
- Dill M, Fahle M. 1998. Limited translation invariance of human visual pattern recognition. *Percept Psychophys*. 60:65-81.
- Drucker DM, Aguirre GK. 2009. Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cereb Cortex*. 19(10):2269-2280.
- Fiser J, Biederman I. 2001. Invariance of long-term visual priming to scale, reflection, translation, and hemisphere. *Vis Res*. 41:221-234.
- Golcu D, Gilbert CD. 2009. Perceptual Learning of Object Shape. *J Neurosci*. 29:13621-13629.
- Goris R, Op de Beeck H. 2009. Neural representations that support invariant object recognition. *Front Comp Neurosci*. 3:3.
- Gross CG, Rocha-Miranda CE, Bender DB. 1972. Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol*. 35:96-111.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. 293:2425-2430.
- Hemond CC, Kanwisher NG, Op de Beeck HP. 2007. A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS one*. 2:e574.
- Hoffman KL, Logothetis NK. 2009. Cortical mechanisms of sensory learning and object recognition. *Philos Trans R Soc Lond*. 364: 321-329.
- Hung CP, Kreiman G, Poggio T, DiCarlo JJ. 2005. Fast readout of object identity from macaque inferior temporal cortex. *Science*. 310:863-866.
- Kravitz DJ, Vinson LD, Baker CI. 2008. How position dependent is visual object recognition? *Trends Cogn Sci*. 12:114-122.
- Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. *Front Syst Neurosci*. 2:4.
- Li N, DiCarlo JJ. 2008. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*. 321: 1502-1507.
- Macevoy SP, Epstein RA. 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol*. 19:943-947.
- Miyashita Y. 1988. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*. 335:817-820.
- Nazir TA, O'Regan JK. 1990. Some results on translation invariance in the human visual system. *Spat Vis*. 5:81-100.
- Niemeier M, Goltz HC, Kuchinad A, Tweed DB, Vilis T. 2005. A contralateral preference in the lateral occipital area: sensory and attentional mechanisms. *Cereb Cortex*. 15:325-331.
- Op De Beeck H, Vogels R. 2000. Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol*. 426:505-518.
- Riesenhuber M, Poggio T. 2000. Models of object recognition. *Nature neuroscience*. 3(Suppl):1199-1204.
- Sayres R, Grill-Spector K. 2008. Relating retinotopic and object-selective responses in human lateral occipital cortex. *J Neurophysiol*. 100:249-267.
- Schwarzlose RF, Baker CI, Kanwisher N. 2005. Separate face and body selectivity on the fusiform gyrus. *J Neurosci*. 25:11055-11059.
- Schwarzlose RF, Swisher JD, Dang S, Kanwisher N. 2008. The distribution of category and location information across object-selective regions in human visual cortex. *Proc Natl Acad Sci U S A*. 105:4447-4452.
- Wallis G, Rolls ET. 1997. Invariant face and object recognition in the visual system. *Prog Neurobiol*. 51:167-194.
- Williams MA, Baker CI, Op de Beeck HP, Shim WM, Dang S, Triantafyllou C, Kanwisher N. 2008. Feedback of visual object information to foveal retinotopic cortex. *Nat Neurosci*. 11:1439-1445.
- Williams MA, Dang S, Kanwisher NG. 2007. Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci*. 10:685-686.
- Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE. 2008. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci*. 11:1352-1360.